View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Asia University Academic Repositories

Identification and Evaluation of Plagiarism amongst Japanese University Students

Michael Schraudner, Asia University

Abstract

Plagiarism is one of many complex issues we face in the second language classroom. For this study, student plagiarism was analyzed over the course of three semesters of Asia University's Freshman English course. Students are in the Freshman English program at Asia University and are required in the class to read seven to ten books and write a short summary about each book over the course of a semester. Data was gathered using Google Forms and tested for plagiarism using a variety of online error correction software programs. Students were told to not copy or plagiarize their assignments at the onset of the semester. Over 50,000 words from 1,400 students' reports were analyzed for this study. Students were also given a questionnaire regarding plagiarism in the class and their feelings on the topic. Their responses both to the questionnaire as well as the actual amount of detected plagiarism in the class will be discussed.

Background

Over the 2012-13 academic year, I began collecting student writing samples and using automated correction programs to analyze the results and to help tailor classes to focus on the highest incidences of mistakes. While completing research for my previous article, I discovered several ways to detect plagiarism using online error correction software. The writing samples consist of the summary section of book reports from students in their first year of intensive English at Asia University in Tokyo, Japan.

This article is part of a longitudinal study using data collected using Google Forms. It is a continuation of a previous study *The Online Teacher's Assistant: Using Automated Correction Programs to Supplement Learning and Lesson Planning* undertaken in the 2012-13 school year. The methodology was identical in this study in that:

> *Using smartphones or computers, students input assignments into an online form, which is then sent to a spreadsheet. The teacher can then easily assess assignments manually as well as by using a variety of automated*

grammar/language tools. While the assignment includes a variety of questions, for this particular data gathering process only one section (Book Summary) was analyzed.

(Schraudner, 2014)

Since the conclusion of the previous study, I began to write and edit Visual Basic macros that could count phrases across a document. Visual Basic is a programming language that (amongst other uses) allows users to do more complex analysis across a variety of Microsoft software including Microsoft Word and Excel. By creating this macro, I was able to input all the students writing samples then scan it as a whole and by class. I could then potentially identify the individual students plagiarizing and other information relating to plagiarism by class.

Capabilities of online plagiarism detection have expanded over the years and the opportunity to test them against the data set will present an interesting way to combat this in the classroom. In many writing classes, Teachers receive hundreds of writing samples per week and checking them or incidents of plagiarism either from the original sources or from among students is incredibly time consuming. This analysis will attempt to identify the method in which the plagiarism was done and ways in which educators can quickly identify the students who do it.

Method

The data pool is comprised of 1,439 book summaries written by Freshman English students in the Intensive English program at Asia University. It contains 52,700 words and was collected over the course of three semesters at the University from April 2012 until August 2014. It consists of three University departments including Law, Economics and Business Hospitality members.

In addition to assessing the data, a questionnaire was given to students of which 55 students responded. The questions referenced plagiarism in Japan and their perceptions of how widespread it is. The questionnaire determined the students' interpretation of the problem of plagiarism and allowed for them to voice their opinions on the topic. There was also a chance to self-report if they had plagiarized in class.

The three main ways plagiarism were checked for in this study were by inputting the data into Grammarly.com's plagiarism detection system, checking for superfluous punctuation and applying a phrase counter to cross reference the data against itself. Grammarly checked the information against online and print sources. According to its website: (Grammarly can) Detect plagiarism by checking your text against over 8 billion web pages (grammarly.com). In order to use both of these tools, students were required to submit their writing samples to Google Forms. Their weekly book reports were then imported to a main spreadsheet, where the data could then be entered into these programs to be examined.

Grammarly.com is an advanced online error correction software that also contains a plagiarism detection component. It is chiefly intended for writers to check their work so that they don't accidentally plagiarize a source and correctly cite their work. When the entire data set is entered into the system however, the corresponding output can be checked against the students' book reports. When the full report is produced, it indicates a percentage of potential plagiarism. The potential plagiarism includes any sentence that the Grammarly program cross-referenced off the internet and has flagged as potentially copied. These incidences are then checked against the student names on the Google Forms database and the student(s) suspected of plagiarism can be identified.

Another method which was employed to check against direct copying from the graded readers was checking the data pool for question marks. Surprisingly when scanning the summary section of the database, students will include all of the punctuation from a certain section that they have copied. Frequently, the books begin with a synopsis of the story and a teaser question. For example :"Will he find the missing items?" or "What will they do next?". This in and of itself is a clear indicator that students have simply taken the material straight from the book. The summary should almost never include a question mark because a summary is not a question, nor should it include one. However, disappointingly the students engaging in copying did not take this into account.

The other way in which plagiarism was checked against was by using a phrase counter. The phrase counter checks the work for identical words/phrases to see if the students have been copying from one another. It can be configured to check for duplicate incidences of clusters of 2 words up to 10 word strings. The program will output a list of

how many times the same phrase is used. Only 10 word strings were checked in this particular study. This is because of the redundant data that would be produced checking for less than 10 word strings (i.e. if the phrases are the same, the counter will report an identical phrase for all 2-10 word strings associated with it).

Clearly one of the ways in which students can easily complete their digitized assignments is by copying and sharing amongst classmates. By using the phrase counter, this method can easily identify pairs (or potentially groups) of students who are copying from one another.

Analysis

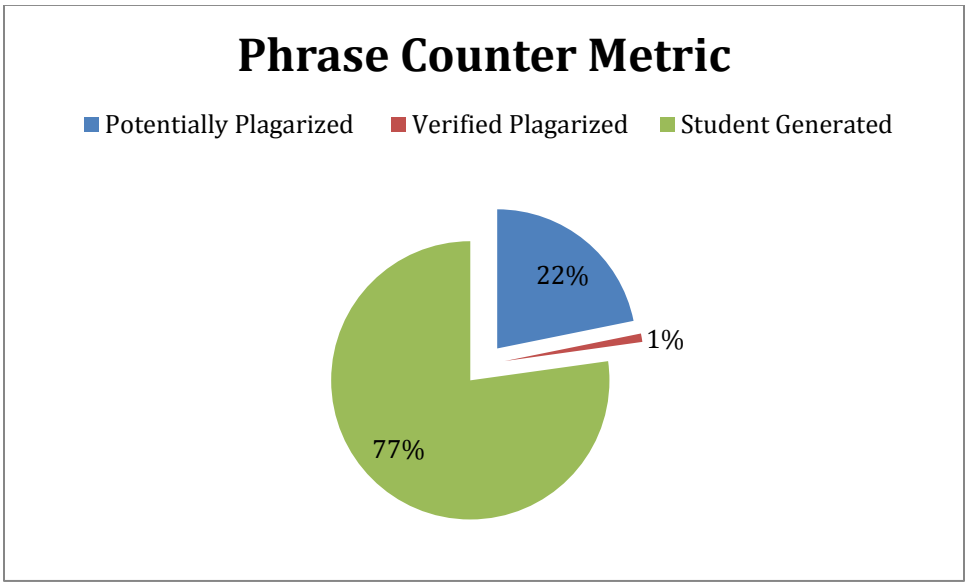
The data was entered into the three plagiarism/copy detection systems beginning with a phrase counter. After trial and error using the phrase counting macro, the macro was set to detect ten word clusters of exact phrases. In the event a potential copying match was discovered through the macro, it was necessary that the potentially plagiarized sentence was cross-referenced against the entire database. This was done by separating the classes into different Microsoft Word documents and using the "Find and Replace" function, where the entire ten word phrase that was detected by the phrase counter macro was searched for across the whole data set. Once the search was completed, the names of the students who were suspected of copying would be displayed in the along with the copied text.

For instances where more than three students had input the same phrase, I have labeled this "verified copying" and when two students had the same phrase entered I have labeled it "potential copying". This is based on the assumption that when three students have entered the exact same information there is simply more evidence to indicate copying has taken place. Please note that the copying could actually indicate plagiarism if multiple students have taken the information from the same book.

In the Business Hospitality section of Freshman English book reports there were 29 potential instances of copying out of 403 Book Reports. That correlates to 27 situations where two students entered the same exact 10 word string of text and two situations where three students entered the same exact 10 word string of text. It is an approximate rate of 7% potential copying with less than .5% verifiable. In the Law section of Freshman English there were 123 potential instances of copying out of 537

Book reports. This is 111 sentences where two students entered the same text and 12 sentences where three or more students entered the same text. This is a rate of approximately 23% potential copying and 2% verifiable. In the Economics section of Freshman English there were 176 potential instances of copying out of 499 book reports. All of these sentences were entered by two students groups so there was no "verified" copying. This is a rate of potential copying of 35% with 0% verified. Overall, averaged out in the three classes there were 328 potential and verifiable instances out of 1,439 reports checked making the average percentage of copying to be approximately 23%.

Figure 1: Phase Counter Metric



The results from the phrase counter showed that copying was in fact taking place on a moderate scale and helped to pinpoint those culpable. It proved to be an indicator of copying, but certainly not a reliable metric to automatically show exactly how many students were plagiarizing or copying. There were several instances where two students had copied off one another, but the phrase counter had counted the same book report sentence multiple times as independent events. Regardless, it was a way to scan through several hundred reports and narrow the field of potentially copied work. The data it provided in my opinion (with the exception of three student copying) was incomplete.

Alternatively, checking the database for question marks yielded some clearer and more actionable information. As mentioned previously, the rationale for checking the

data set for question marks is that they should only be present if a student has either plagiarized directly from the text or copied plagiarized text from another student. They should not exist within a book summary. Question marks occasionally occurred twice in the same book report. The findings were adjusted just to show if they existed at all in the book report as to indicate copying. The question marks were discovered by importing the individual class data to Microsoft Word, running the "Find" search and inputting a question mark.

Business Hospitality had five question marks in their 403 book reports which potentially indicated at least five students copied directly from the book for 1% of all book reports. Economics students had 13 question marks in 499 reports for 2.6% of all reports handed in. The Law students had 28 question marks in 537 book reports for 5.2%. By the question mark metric, 46 question marks were found making the total percentage for all 1,439 Book Reports to be 3.1% copied.

Finally, the information was entered into the grammarly.com plagiarism detection software. The writing samples were copied from the database and entered by specific class into the website. Because the Grammarly website is, by its own definition, a tool for writers, it approaches plagiarism as "percentage of borrowed text that may require citation."(Grammarly, 2014). The results for "potential plagiarism" were as follows: Business Hospitality 19% potentially plagiarized Economics 35% and Law 29%. The average amount of plagiarism according to Grammarly's plagiarism checker, across all classes, was 28%.

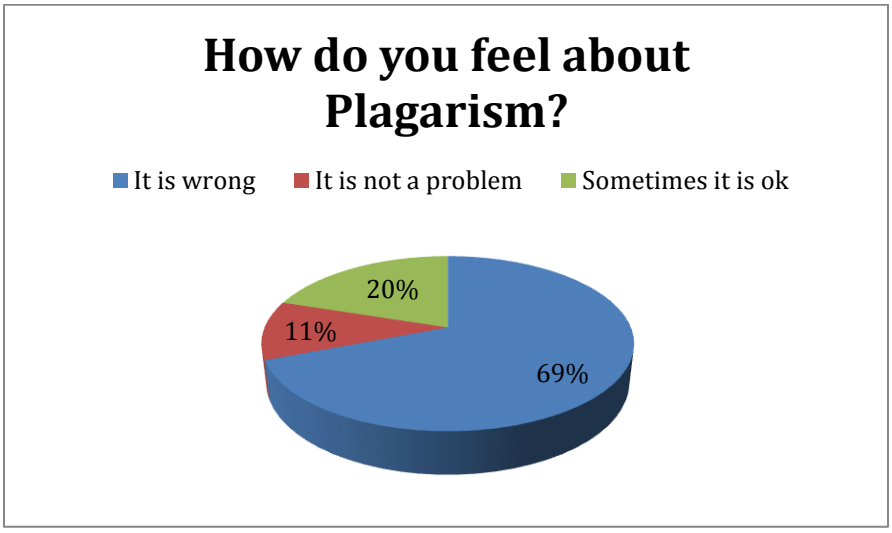
The plagiarism percentage figures from grammarly.com when cross referenced with the phrase counter create the plagiarism margin of difference (PMD) and were fairly consistent with the exception of the Business Hospitality class. There was a 6% difference between the Law classes' PMD (29% Grammarly, 23% Phrase Counter). Business Hospitality students' PMD was 12% (19% Grammarly but 7% phrase counter). Surprisingly, there was no difference for the Economics students' PMD both measured out the highest of all the classes at 35%. The total average plagiarism detected according to Grammarly was approximately 28% and 23% with the phrase checker making the total PMD 5%.

In the future, I would like to create a more advanced detection program that uses all three of these metrics to check within a database without having to use multiple software platforms. The calculations involved in determining these percentages can be streamlined into a database in an easier, more accessible way.

Survey

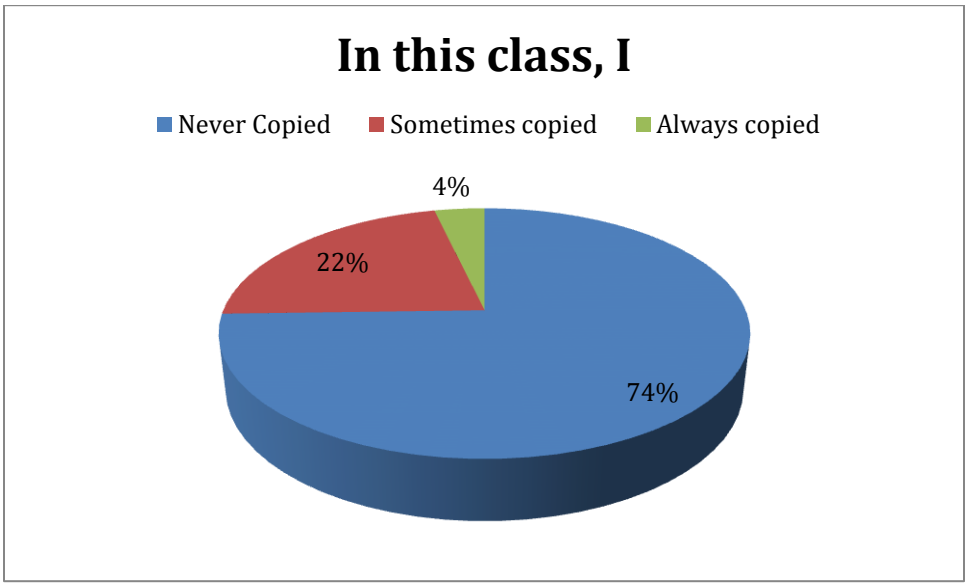
Having spent a considerable amount of time thinking of methods to discover plagiarism in the classroom, I thought asking the students their opinions about the issue would yield interesting results. After advising them at the beginning of the semester that plagiarism and copying were actively discouraged in the classroom, what were there actual opinions? Would they admit to copying from their books? This anonymous survey was given to the students during the last week of classes of the Spring semester in 2014:

Figure 2: Question 1



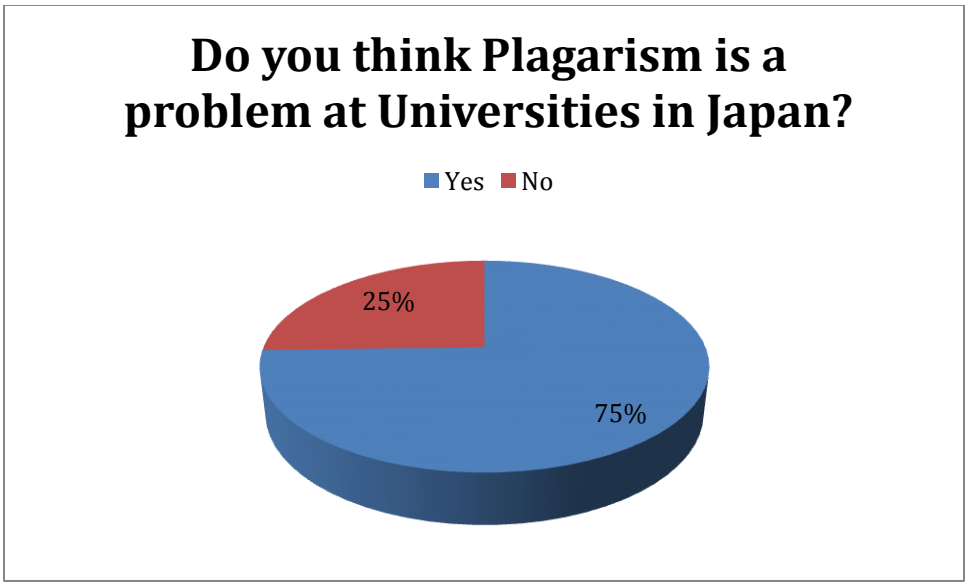
The vast majority of students, 69% said that plagiarism is wrong, while surprisingly 11% said it is not a problem and 20% said it was sometimes ok. This brings up several questions. If 31% of students expressed that plagiarism or copying is of little consequence to them, and with no formal policy given by the department, how should the teacher react to instances of it in the classroom? How should it be checked for and/or punished?

Figure 3: Question 2



The students' honesty in this question surprised me quite a bit. 74% said they didn't copy while 22% said they sometimes copied and 4% said they always copied. This fits almost exactly with the data gathered from the plagiarism detection programs with a very small margin of error. 26% of students self-reported that they at least occasionally took part in copying/plagiarizing.

Figure 4: Question 3



Overwhelmingly, 75% of students believed that plagiarism is a problem at universities in Japan. The remaining 25% that replied in the negative was not far off from the 26% of students that self-reported copying.

Conclusion

Plagiarism is clearly outside the bounds of academic integrity that we as teachers strive to uphold. Interestingly, there is no official discussion of plagiarism or copying within the Asia University Center of Language Education Official Handbook. It is expected that teachers handle this on a case-by-case basis and the guidelines are not expressly conveyed as such.

However, plagiarism in language acquisition is not as cut and dry as in other areas of study. Students are frequently expected to memorize and generate predetermined sentences and answers. Early stages of language are often dependent on recitation and repetition of basic elements. This brings up the issue of how stringently to check for plagiarism or copying in student work because of how beneficial memorization and repetition are in the ESL learner's tool box. Researcher Larry Ellis writes as his top *Principle of Instructed Language Learning (2005)* that: "Language teaching should ensure that learners develop both a rich repertoire of formulaic expressions and a rule-based competence", (p. 33). If this is the case, then perhaps plagiarism can almost be seen as a benefit to the students as opposed to other disciplines where it is considered to be disingenuous. Students are still processing and acquiring their "formulaic expression" repertoire and it may not be in their best interest to discourage them from repeating or copying native speakers.

On the other hand, by its very nature it is passing off others work as one's own writing. The idea of ownership and authorship in work within most educational and literary communities is strictly enforced. To give language students a free pass to copy could be seen as disrespectful to other departments' authenticity and a black mark on language teachers for allowing it to take place.

It could be argued that the assistance of copying within the context of a homework assignment is not correct but better than trying to struggle to create form. Clearly, it is beneficial for students in all stages of language acquisition to have

comprehensible input, but should their output be so stringently monitored? In his study of students at Nanjing University, Yanren Ding concluded that :

> "Successful learners are often seen as having exceptional aptitudes…Their success came from years of practice in imitation, memorization and communication, which was usually first forced upon them by their teacher, but later came to be driven by motivation arising from initial success, teacher praise and personal interest. (Ding, 2007).

The issue lies in the interpretation of the instructor and the institution as to where the boundary occurs. With an internet-based assignment, it could be argued that many students are simply "cutting and pasting" and that this does not benefit the student in terms of enhancing their formulaic expression building. In my experience, the most disappointing aspect of student copying is when students copy from their classmate's incorrect work and reinforce mistakes. Especially within the context of this particular study, it is entirely possible that students copied sections of the text from the books (or each other) and entered them in as their summary. But then, when is it that plagiarism occurred? After the 4th repeated word? The 5th? Is it not in actuality assisting their English to be sitting with the book, scanning for meaning and entering a phrase that is linguistically correct yet borrowed from the source material?

All students were reprimanded after the results of this study became public to them and were reminded again that any form of academic dishonesty including cheating is expressly forbidden in class. Students that were caught copying from one another, or plagiarizing entire portions of their book reports were stripped of credit for the assignments. Although copying and rote memorization has its place in the language classroom, it is in my opinion that it should not take place during a free writing activity. There are other opportunities that students have to reproduce form. It is important to remember that we are acting as the gatekeepers of language, and that we are trying to instill an appreciation and motivation in our students to not cut corners.

The data gathered in this study confirms that plagiarism is taking place but the exact scale of it is difficult to exactly measure using only these particular automated programs. It is important to note that the percentage of plagiarism, reflects only the amount of plagiarized sentences and not the number of individual students who

plagiarized. In fact it was possible that only a small portion of the students were copying/plagiarizing and the vast majority were reading and generating original reports.

The numbers from this study may presenting a more dismal situation than is actually occurring. In order to improve it, I will have to create a more complex macro (or program) that displays all the information in an easier to read format. My goal is to be able to catch plagiarism immediately and have the automated program quickly send an email to both teacher and student after it runs a plagiarism check. Going forward I hope these programs can help teachers become more involved in their classes and encourage students to work to their full potential.

References

Asia University CELE handbook. (2014). Tokyo, Japan: Asia University Press.

Ding, Y. (2007). Text memorization and imitation: The practices of successful Chinese learners of English, *System, 35*(2), 271–280.

Ellis, R. (2005). Principles of instructed language learning. *System, 33*(2), 209-224.

Schraudner, M. (2014). The online teacher's assistant: using automated correction programs to supplement learning and lesson planning. *CELE Journal*. *22*, 128-40.

Appendix

Plagiarism Questionnaire

Plagiarism Questionnaire: Please write your answer to these questions. Please do not write your name

How do you feel about plagiarism

___

a. It's wrong
b. Its not a problem
c. Sometimes its ok

In this class, I:

___

a. Never copied
b. Sometimes copied
c. Always copied

Do you think plagiarism is a problem at Universities in Japan?

___

a. Yes
b. No

If another student was plagiarizing in class, I would:

___

a. Say nothing
b. Tell them not to do it
c. Tell the teacher

Tell me your thoughts about copying/plagiarism: